



Programación Python para BigData

Lección 8: Apache Spark con PySpark [2/2]

Spark MLib empleando Iris dataset

ACTIVIDAD LECCIÓN 8

Objetivos

- | Realizar la predicción del Iris Dataset usando Spark MLib

Contenido correspondiente a lección 8:

1. Spark MLib.
2. Spark GraphX.
3. Diferencia entre Hadoop y Spark

Actividad relacionada con la lección 8:

El/la alumno/a deberá entregar un archivo .ipynb como actividad con las siguientes partes:

- | Transformar en vector los datos numéricos usando VectorAssembler: **2,0 ptos**
- | Transformar los datos correspondientes a la columna especies usando StringIndexer: **2,0 ptos**
- | Calcular la predicción y la precisión del modelo empleando Decision Tree Classifier : **2,0 ptos**
- | Calcular la predicción y la precisión del modelo empleando Gradient-boosted tree classifier: **2,0 ptos**
- | Calcular la predicción y la precisión del modelo empleando Random Forest Classifier : **2,0 ptos**



Nota: Adjuntar junto con la actividad el dataset empleado, poner como nombre de la actividad **nombre_apellido_spark2.ipynb** y explicar detalladamente cada paso.

NOTA FINAL: 10 PTOS